Market Prices of Houses in Atlanta

Eric Slone, Haitian Sun, Po-Hsiang Wang

ECON 3161 – Prof. Dhongde

Abstract

This study reveals the relationships between residential property asking price and various explanatory variables, including home square footage, land parcel square footage, number of bedrooms, and year of construction, in Atlanta, Georgia. Simple and multiple linear regressions were developed using the method of ordinary least squares. Square footage of the house was found to be the most influential factor in determining asking price, number of bathrooms exhibited a significant relationship, and garage capacity and year of construction were found to be the weakest relationships.

I. Introduction

One of the most diverse real estate markets in the country, Atlanta housing prices exhibit large fluctuations across geographic locations. Many factors affect residential housing prices across the metropolitan area, including land parcel size, home square footage, and amenities. Intuition leads to the conclusion land parcel size will have higher impacts on residential housing prices towards the city center, while suburban area prices will be more affected by home square footage and amenities, such as garages, swimming pools, and fireplaces.

The objective of this study is to reveal the relationships between the dependent variable, asking price, with the explanatory variables, including land parcel size, home square footage, number of bedrooms and bathrooms, garage capacity, and neighborhood. Using the ordinary least squares regression method, the constructed model will enable the prediction of housing market values based on various property characteristics, thus enabling buyers and sellers alike to create a preliminary appraisal of any property in Atlanta.

Data were gathered on April 15th, 2014. The city of Atlanta was divided into five distinct neighborhoods: Intown, Southeastern, Southwestern, Northwestern, and Buckhead. Twenty single family houses were randomly selected from each neighborhood, resulting in 100 observations.

II. Literature Review

Hanonnen (2008) investigates the best practices and modeling techniques when attempting to find relationships between various variables, including permitted building volume, house price index, northing, and easting. An ordinary least squares estimation resulted in all explanatory variables being plausible in sign and magnitude, indicating the OLS method is accurate. Other modeling methods included robust MM-estimation, structural time-series estimation, and robust local regression. While permitted building volume, house price index, northing, and easting all had significant relationships with market values, land parcel size was found to have a much weaker relationship. The data used for the study was gathered in the city of Espoo, Finland.

Inspired by the significant increase in United States legislative action in favor of land conservation during the 1994-2005 period, Cho (2009) examines the relationship between both land parcel size and proximity to open space with housing sale prices. Particularly, the study investigated the trade-off between the two

explanatory variables, and found that proximity to open spaces, such as parks, greenways, and bodies of water, is a very close substitute for land parcel size in relation to sale prices. All of the data used in the study was obtained from a single county in Tennessee.

Abelson (2013) models median housing prices in various neighborhoods across Sydney, Australia using numerous explanatory variables, including land parcel size, housing size, access to public transportation, and housing density. The findings include significant relationships between sale price and many of the explanatory variables, especially land parcel and housing sizes.

Cebula (2009) analyzed 2,888 single-family homes in the city of Savannah, Georgia. Using the method of ordinary least squares, the models constructed in the study revealed positive relationships between the dependent variable, housing price, and multiple explanatory variables, including number of bedrooms, bathrooms, fireplaces, stories, garage capacity, and various amenities. Spatial relationships were also analyzed; proximity to open land, lakes, and the presence of cul-de-sacs were found to positively affect home prices, while proximity to apartment complexes and busy roads negatively affected sale prices.

Although no literature could be located analyzing the factors of housing sale prices in the Atlanta, Georgia area, all of the aforementioned studies have relevant findings in other areas. Hanonnen (2008) provided insight into various modeling techniques, and provided assurance the OLS method, which is used in this study, is best for the purposes of discovering relationships between various explanatory variables and housing prices. Abelson (2013) provided relevant insight into the explanatory variables affecting housing prices, which can be compared to the findings of housing price factors in Atlanta, Georgia, analyzed in this study. The most relevant literature was found in The Hedonic Pricing Model to the Housing Market of the City of Savannah, in which various explanatory variables, many identical to those analyzed in this study, were found to both positively and negatively affect home prices in the city of Savannah.

III. Data

The data used in this study were collected from www.realtor.com. Residential lot asking price and home characteristics were obtained for properties in five Atlanta city neighborhoods: Intown, Southeastern, Southwestern, Northwestern, and Buckhead. Twenty single family homes were randomly selected from each neighborhood, resulting in 100 observations.

The property asking price represents the dependent variable, and five quantitative independent variables are used in the OLS regression: home square footage, lot square footage, number of bedrooms, number of bathrooms, year of construction, and number of cars accommodated in a garage. The primary Atlanta neighborhoods represent the qualitative variables, and Intown Atlanta was used as the base neighborhood.

The data collected for this study are used to specify the model, create parameter estimates, check for model adequacy, reiterate until the model is adequate, validate the model, reiterate until the model is valid, and then use the final model to estimate asking price of property within each Atlanta neighborhood. Table 1 below outlines all of the variables used in this study.

Table 1. Variables used in OLS Regression

Dependent Variable	**Independent Variables**
Asking Price (USD)	Home Square Footage (ft^2)
	Lot Square Footage (ft^2)
	Number of Bedrooms
	Number of Bathrooms
	Year Built
	Number of Cars Garage Accommodates
	Neighborhood (qualitative)

An analysis of the data reveals that all of the Gauss-Markov assumptions are satisfied. The model utilizes linear regression, so the first assumption of linear parameters is satisfied. All of the data were collected randomly from large lists of housing characteristics, so the assumption of random sampling is satisfied. Due to the impossibility of perfect correlation between any of the explanatory variables, the third assumption is satisfied. The highest correlation was found to be between the number of bathrooms and home square footage, with a value of 0.81. Lastly, all of the relevant information available for each home was utilized, so the model is not believed to be under- or mis-specified, and the assumption of zero conditional mean is satisfied. Table 2 below outlines the summary statistics for each of the variables used in this study.

Table 2. Variable Summary Statistics

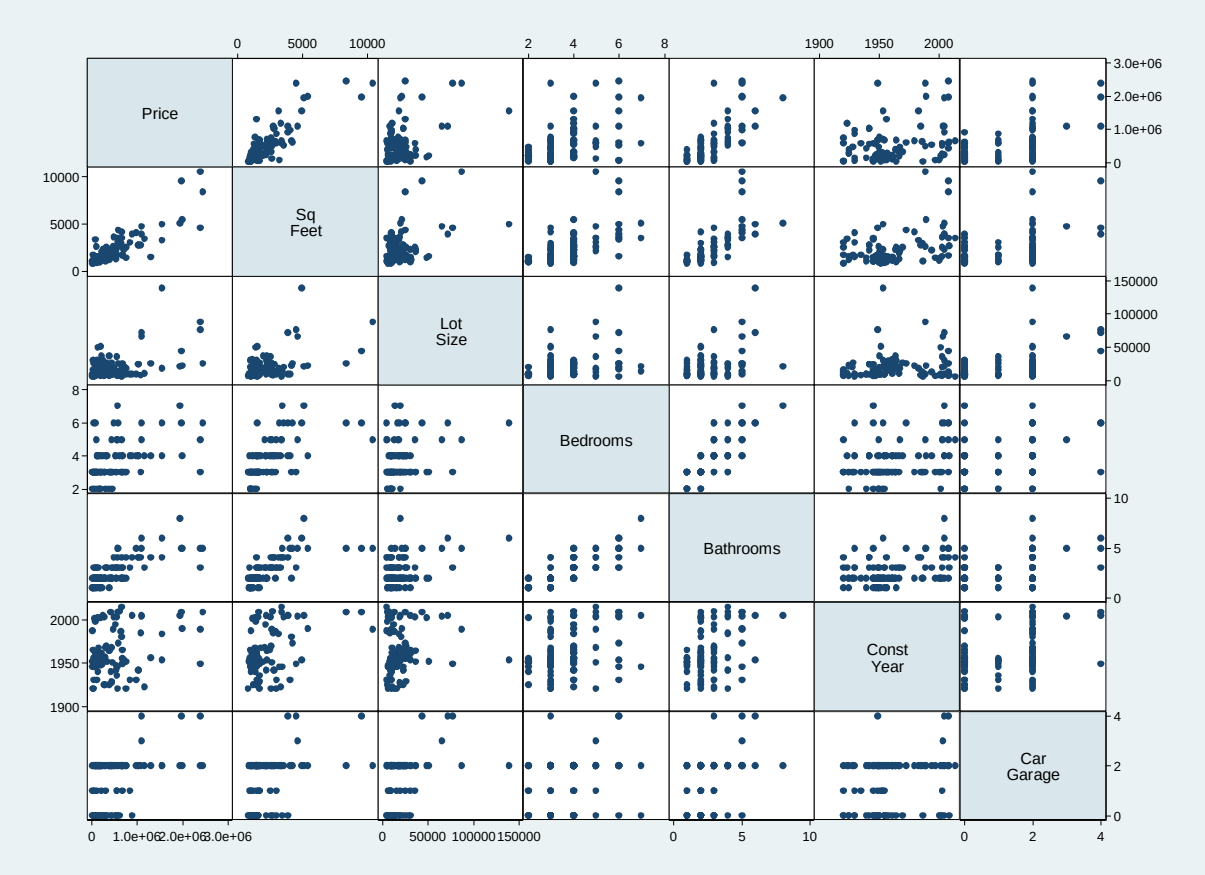
Variable	Obs	Mean	Std. Dev.	Min	Max
price	100	517776.4	550770.3	12100	2450000
bedrooms	100	3.63	1.142874	2	7
bathrooms	100	2.64	1.322074	1	8
sqfeet	100	2302.34	1673.147	845	10472
lotsize	100	20213.59	19366.73	4520	139392
constyear	100	1961.76	25.94093	1920	2014
cargarage	100	1.37	1.021536	0	4

Table 3 below outlines the correlations between each explanatory variable. All of the variables are positively correlated, as expected. Some variables are highly correlated. Bedrooms and sqfeet, as well as bathrooms and bedrooms have correlation values of 0.8. This indicates that there is a possibility of multicollinearity, so several different models need to be tested. In Table 4, all of the explanatory variables present positive relationships with price since the slopes seem to be positive, consistent with positive correlations.

Table 3. Explanatory Variable Correlation

	lnPrice	lnSqFeet	lnLotS~e	bedrooms	bathro~s	lnCons~r	cargar~e
lnPrice	1.0000						
lnSqFeet	0.7646	1.0000					
lnLotSize	0.3857	0.4262	1.0000				
bedrooms	0.4092	0.6948	0.2983	1.0000			
bathrooms	0.6726	0.8077	0.3411	0.8068	1.0000		
lnConstYear	0.2583	0.4032	0.1471	0.3156	0.3777	1.0000	
cargarage	0.4111	0.4613	0.4249	0.2309	0.3464	0.2376	1.0000

Table 4. Scatter Plot



III. Results

The collected data were first used to develop a simple linear regression using the method of ordinary least squares. Home asking price was taken to be the dependent variable, and home square footage was used for the explanatory variable. Table 6 below outlines the results of the simple linear regression.

Table 6. Simple Linear Regression Results

Source	SS	df	MS	Number of obs =	100
				F(1, 98) =	137.90
Model	80.1794221	1	80.1794221	Prob > F =	0.0000
Residual	56.9793399	98	.581421835	R-squared =	0.5846
				Adj R-squared =	0.5803
Total	137.158762	99	1.38544204	Root MSE =	.76251

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnSqFeet	1.646853	.1402391	11.74	0.000	1.368553	1.925153
_cons	.1233631	1.064353	0.12	0.908	-1.988811	2.235537

The results above correspond with intuition; the positive coefficient indicates the positive relationship between asking price and home square footage. A percent change of square footage results in a 1.64% change in the asking price. The p-value is 0.000, which is less than 0.01; therefore, this variable is statistically significant at 1%. Considering only one explanatory variable was used in this regression, the r-squared value of 0.58 is much higher than expected.

The data were then used to develop a multiple linear regression using the method of ordinary least squares. Table 7 below outlines the results.

Table 7. Multiple Linear Regression Results

Source	SS	df	MS		
Model	117.255603	10	11.7255603	Number of obs =	100
Residual	19.9031586	89	.223630995	F(10, 89) =	52.43
				Prob > F =	0.0000
Total	137.158762	99	1.38544204	R-squared =	0.8549
				Adj R-squared =	0.8386
				Root MSE =	.4729

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnSqFeet	.5617323	.1909511	2.94	0.004	.1823166	.941148
lnLotSize	.2775955	.0839554	3.31	0.001	.1107778	.4444131
bedrooms	-.1599707	.0803623	-1.99	0.050	-.3196488	-.0002926
bathrooms	.2256127	.0796534	2.83	0.006	.0673431	.3838823
lnConstYear	-.5590084	4.202085	-0.13	0.894	-8.908461	7.790444
cargarage	.0336868	.0575052	0.59	0.559	-.0805749	.1479484
southeastern	-1.122568	.1789034	-6.27	0.000	-1.478045	-.7670905
southwestern	-1.736762	.1810708	-9.59	0.000	-2.096545	-1.376978
northwestern	-.6161533	.1762393	-3.50	0.001	-.966337	-.2659696
buckhead	.0047453	.1674597	0.03	0.977	-.3279936	.3374843
_cons	10.53205	31.65188	0.33	0.740	-52.35957	73.42367

The results above once again correspond with intuition; home square footage, land parcel size, number of bathrooms, and garage capacity all have a positive effect on the asking price. The negative relationship between the year of home construction and the market price is not as expected; however, possible explanations could include the fact older homes often feature better construction, unique architectural features, and larger living areas. The number of bedrooms has a negative relationship with the market price, and is also another counter-intuitive variable. Two simple linear regression models were tested with these two variables. Each of the models resulted in a positive coefficient. Therefore, the negative coefficient may be due to effects on the model from other explanatory variables. The home square footage is the main factor affecting the dependent variable; a percent increase in lot square footage results in a 0.56% change in the asking price. lnConstYear, Cargarage, and buckhead have very high p-values. These variables are not statistically significant. Bedrooms is significant at 5%, while all other variables are significant at 1%. It is important to note the home square footage coefficient decreased as compared to the simple linear regression result, indicating a bias existed due to the omission of important explanatory variable in the simple regression. The adjusted r-squared value of 0.84 is exceptionally high.

Holding all other factors constant, the location of a home within the Atlanta city limits can significantly affect the home market price. Buckhead proved to have the most positive affect on home prices, a conclusion consistent with local opinion and prestige, while the Southwestern area of Atlanta had the most negative affect, consistent with the high crime and poverty rates characterizing the region. The Intown neighborhood, which includes the Downtown and Midtown skyscraper regions, was used as the base neighborhood value.

To test whether other neighborhoods are statistically different from each other, models using different neighborhoods as the base levels were tested. In Appendix, the results from those regression models indicate that all of the five areas are significantly different from each other except for Intown and Buckhead, since the p-value of those variables are much smaller than 0.01. The null hypothesis of the areas having the same price is therefore rejected. The lowest price to highest price region is given in the following order: southwestern, southeastern, northwestern, Intown and Buckhead.

Variables that are not significant were removed from the model in an attempt to improve the regression model. Table 8 below outlines the results

Table 8. Multiple Linear Regression Results

Source	SS	df	MS		
Model	117.170841	7	16.7386916	Number of obs	= 100
Residual	19.987921	92	.21726001	F(7, 92)	= 77.04
				Prob > F	= 0.0000
Total	137.158762	99	1.38544204	R-squared	= 0.8543
				Adj R-squared	= 0.8432
				Root MSE	= .46611

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnSqFeet	.586928	.173153	3.39	0.001	.2430312	.9308249
lnLotSize	.2916624	.0788958	3.70	0.000	.1349686	.4483562
bedrooms	-.1669782	.0775984	-2.15	0.034	-.3210952	-.0128612
bathrooms	.2271005	.0770796	2.95	0.004	.0740137	.3801873
southeastern	-1.11691	.1559541	-7.16	0.000	-1.426648	-.8071716
southwestern	-1.740296	.1662889	-10.47	0.000	-2.07056	-1.410032
northwestern	-.6150128	.1455541	-4.23	0.000	-.9040958	-.3259298
_cons	6.035489	1.158854	5.21	0.000	3.733905	8.337073

The interpretations of the coefficients are identical to the previous model. However, these variables became more significant as the insignificant variables were eliminated from the model. R-squared decreases as expected but only a very slightly decrease from 0.8549 to 0.8543. However, R-square Adj increased from 0.8386 to 0.8432, suggesting a more accurate model.

To compare the two multiple regression model, an F- test was conducted with the null hypothesis defined as joint significance exists in the model. The F statistic was calculated to be 0.19, which is much smaller than 2.71, the critical value of F-distribution at 5% with (3, 92) degrees of freedom. The null hypothesis cannot be rejected, suggesting the model is more accurate without the extra variables.

Table 9 below outlines the results from all three models.

Table 9. Results

Dependent Variable SellingPrice			
Independent Variables	**Single Regression**	**Multiple Regression**	**Final Multiple Regression**
lnsqfeet	1.64*** (11.74)	0.561*** (2.94)	0.587*** (3.39)
lnsqftlot	-	0.277*** (3.31)	0.292*** (3.70)
bedroom	-	-0.160* (-1.99)	-0.167** (-2.15)
bathroom	-	0.226*** (2.83)	0.227*** (2.95)
lnconstyear	-	-0.559 (-0.13)	-
cargarage	-	0.034 (0.59)	-
southeastern	-	-1.123*** (-6.27)	-1.12*** (-7.16)
southwestern	-	-1.737*** (-9.59)	-1.740*** (-10.47)
northwestern	-	-0.616*** (-3.5)	-0.615*** (-4.23)
buckhead	-	0.005 (0.03)	-
Intercept	0.123 (0.12)	10.532 (0.33)	6.035*** (5.21)
No. of obs.	100	100	100
R-square	0.5846	0.8549	0.8543
R-square Adj.	0.5803	0.8386	0.8432

*** significant at 1% ** significant at 5% *significant at 10%

IV. Conclusions

In this study, the relationships between various home characteristics and the asking price of a residential property were analyzed using both a simple linear regression and a multiple linear regression using the method of ordinary least squares. Home square footage was utilized as the explanatory variable in the simple linear regression, and the multiple linear regression consisted of the addition of land parcel size, number of bedrooms, year of construction, and other explanatory variables. The results of the multiple linear regression proved the bias due to the omission of important factors in the simple linear regression. Home square footage was found to be the most important factor in the determination of residential property price, while garage capacity proved to be the weakest factor. Home location also significantly affects market prices within the Atlanta city limits; Buckhead proved to have the most positive effect on home values, while Southwestern Atlanta had the most negative effect. The exceptionally high adjusted r-squared value of 0.84 indicates the model constructed in this study is valid for estimating the asking price of a residential property using various home characteristics.

References

Abelson, Peter, Roselyne Joyeux, and Stephane Mahuteau. "Modelling House Prices Across Sydney." *Australian Economic Review*. 46.3 (2013): 269-285. Print.

Cebula, Richard. "The Hedonic Pricing Model Applied to the Housing Market of the City of Savannah." *Review of Regional Studies* 39 (): 9-22. Print.

Cho, Seong-Hoon, Christopher Clark, William Park, and Seung Kim. "Spatial and Temporal Variation in the Housing Market Values of Lot Size and Open Space." *Land Economics*. (2009): n. page. Print.

Hanonnen, Marko. "Predicting Urban Land Prices: A Comparison of Four Approaches." *International Journal of Strategic Property Management*. 12. (2008): 217-236. Print.

"Homes for Sale." *Realtor*. National Association of Realtors, 05 03 2014. Web. 12 Mar 2014. <http://www.realtor.com>.

Appendix

Source	SS	df	MS		
Model	117.255603	10	11.7255603	Number of obs	= 100
Residual	19.9031586	89	.223630995	F(10, 89)	= 52.43
				Prob > F	= 0.0000
Total	137.158762	99	1.38544204	R-squared	= 0.8549
				Adj R-squared	= 0.8386
				Root MSE	= .4729

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnSqFeet	.5617323	.1909511	2.94	0.004	.1823166	.941148
lnLotSize	.2775955	.0839554	3.31	0.001	.1107778	.4444131
bedrooms	-.1599707	.0803623	-1.99	0.050	-.3196488	-.0002926
bathrooms	.2256127	.0796534	2.83	0.006	.0673431	.3838823
lnConstYear	-.5590084	4.202085	-0.13	0.894	-8.908461	7.790444
cargarage	.0336868	.0575052	0.59	0.559	-.0805749	.1479484
southwestern	-.614194	.1527023	-4.02	0.000	-.9176102	-.3107777
northwestern	.5064144	.1560292	3.25	0.002	.1963877	.8164411
buckhead	1.127313	.1897223	5.94	0.000	.7503388	1.504287
intown	1.122568	.1789034	6.27	0.000	.7670905	1.478045
_cons	9.409478	31.69877	0.30	0.767	-53.5753	72.39425

Source	SS	df	MS		
Model	117.255603	10	11.7255603	Number of obs	= 100
Residual	19.9031586	89	.223630995	F(10, 89)	= 52.43
				Prob > F	= 0.0000
Total	137.158762	99	1.38544204	R-squared	= 0.8549
				Adj R-squared	= 0.8386
				Root MSE	= .4729

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnSqFeet	.5617323	.1909511	2.94	0.004	.1823166	.941148
lnLotSize	.2775955	.0839554	3.31	0.001	.1107778	.4444131
bedrooms	-.1599707	.0803623	-1.99	0.050	-.3196488	-.0002926
bathrooms	.2256127	.0796534	2.83	0.006	.0673431	.3838823
lnConstYear	-.5590084	4.202085	-0.13	0.894	-8.908461	7.790444
cargarage	.0336868	.0575052	0.59	0.559	-.0805749	.1479484
southeastern	.614194	.1527023	4.02	0.000	.3107777	.9176102
northwestern	1.120608	.1619436	6.92	0.000	.7988299	1.442387
buckhead	1.741507	.2014707	8.64	0.000	1.341189	2.141825
intown	1.736762	.1810708	9.59	0.000	1.376978	2.096545
_cons	8.795285	31.68553	0.28	0.782	-54.16319	71.75376

Source	SS	df	MS	Number of obs =	100
				F(10, 89) =	52.43
Model	117.255603	10	11.7255603	Prob > F =	0.0000
Residual	19.9031586	89	.223630995	R-squared =	0.8549
				Adj R-squared =	0.8386
Total	137.158762	99	1.38544204	Root MSE =	.4729

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnSqFeet	.5617323	.1909511	2.94	0.004	.1823166	.941148
lnLotSize	.2775955	.0839554	3.31	0.001	.1107778	.4444131
bedrooms	-.1599707	.0803623	-1.99	0.050	-.3196488	-.0002926
bathrooms	.2256127	.0796534	2.83	0.006	.0673431	.3838823
lnConstYear	-.5590084	4.202085	-0.13	0.894	-8.908461	7.790444
cargarage	.0336868	.0575052	0.59	0.559	-.0805749	.1479484
southeastern	-.5064144	.1560292	-3.25	0.002	-.8164411	-.1963877
southwestern	-1.120608	.1619436	-6.92	0.000	-1.442387	-.7988299
buckhead	.6208986	.1744298	3.56	0.001	.2743104	.9674869
intown	.6161533	.1762393	3.50	0.001	.2659696	.966337
_cons	9.915893	31.70622	0.31	0.755	-53.0837	72.91548

Source	SS	df	MS	Number of obs =	100
				F(10, 89) =	52.43
Model	117.255603	10	11.7255603	Prob > F =	0.0000
Residual	19.9031586	89	.223630995	R-squared =	0.8549
				Adj R-squared =	0.8386
Total	137.158762	99	1.38544204	Root MSE =	.4729

lnPrice	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lnSqFeet	.5617323	.1909511	2.94	0.004	.1823166	.941148
lnLotSize	.2775955	.0839554	3.31	0.001	.1107778	.4444131
bedrooms	-.1599707	.0803623	-1.99	0.050	-.3196488	-.0002926
bathrooms	.2256127	.0796534	2.83	0.006	.0673431	.3838823
lnConstYear	-.5590084	4.202085	-0.13	0.894	-8.908461	7.790444
cargarage	.0336868	.0575052	0.59	0.559	-.0805749	.1479484
southeastern	-1.127313	.1897223	-5.94	0.000	-1.504287	-.7503388
southwestern	-1.741507	.2014707	-8.64	0.000	-2.141825	-1.341189
northwestern	-.6208986	.1744298	-3.56	0.001	-.9674869	-.2743104
intown	-.0047453	.1674597	-0.03	0.977	-.3374843	.3279936
_cons	10.53679	31.68034	0.33	0.740	-52.41136	73.48495